

I know that “Kiki” is angular:
The metacognition underlying sound-shape correspondences

Yi-Chuan Chen^{1,2}, Pi-Chun Huang³, Andy Woods¹, & Charles Spence¹

¹ Department of Medicine, Mackay Medical College, New Taipei City, Taiwan

² Department of Experimental Psychology, University of Oxford, Oxford, UK

³ Department of Psychology, National Cheng Kung University, Tainan, Taiwan

Running head: Metacognition & The Bouba/Kiki Effect

Word count: 3949 words

Submitted to: *Psychonomic Bulletin & Review* (Brief Report)

Date: 17 Jun 2018

*Corresponding:

Dr. Yi-Chuan Chen

Department of Medicine, Mackay Medical College

No.46,Sec.3, Zhongzheng Rd., New Taipei City, Sanzhi Dist., 252, Taiwan

Tel: +886 (2) 26360303 #1255

Fax: +886 (2) 26361295

Email: ycchen@mmc.edu.tw

Abstract

We examined the ability of people to evaluate their confidence when making perceptual judgments concerning a classic crossmodal correspondence, *the Bouba/Kiki effect*: People typically match the “Bouba” sound to more rounded patterns while matching the “Kiki” sound to more angular patterns instead. For each visual pattern, individual participants were more confident about their own matching judgments when they happened to fall in line with the consensual response regarding whether the pattern was rated as “Bouba” or “Kiki”. Logit-regression analyses demonstrate that participants’ confidence ratings and their matching judgments were predictable by similar regression functions. This implies that the consensus and confidence underlying the Bouba/Kiki effect are underpinned by a common process whereby perceptual features in the patterns are extracted and then used to match the sound following rules of crossmodal correspondences. Combining both matching and confidence measures allows one to explore and quantify the strength of associations in human knowledge.

Keywords: Bouba/Kiki effect; Crossmodal correspondences; The consensuality principle; Confidence rating; Radial frequency patterns

Introduction

Most environmental events generate signals that are detectable by multiple sensory modalities, and these signals consist of features corresponding to each other rather than being arbitrarily related (see Spence, 2011). In turn, crossmodal correspondences constitute critical cues for the human brain when inferring whether multiple sensory signals should be integrated or segregated (e.g., Chen & Spence, 2017). One of the best-studied crossmodal correspondences concerns people's matching of the nonsense speech sound "Bouba" to more rounded shapes while matching the nonsense speech sound "Kiki" to shapes that are more angular instead. There are no objectively correct answers to such matchings, only more or less consensual ones. The Bouba/Kiki effect was first reported almost 90 years ago (Holland & Wertheimer, 1964; Köhler, 1929, 1947), and has been reliably replicated in both infants and toddlers (Maurer, Pathman, & Mondloch, 2006; Ozturk, Krehm, & Vouloumanos, 2013), as well as in some populations that are remote from Western culture (e.g., Bremner et al., 2013; Davis, 1961). Hence, the Bouba/Kiki effect is highly reliable even when testing people encountering these stimuli for the first time. In the present study, beyond the well-documented interpersonal consensus, we investigated whether people can monitor their internal process when judging whether a particular sound and shape go "better" together than other pairs. That is, we were interested in the metacognitive component of the Bouba/Kiki effect (e.g., Deroy, Spence, & Noppeney, 2016; Faivre, Filevich, Solovey, Kühn, & Blanke, 2018).

The rudimentary forms of the Bouba/Kiki effect observed in preverbal infants (Ozturk et al., 2013) suggest that this sound-shape correspondence does not solely result from language learning (e.g., the resemblance between rounded/angular visual shapes and lip movements when uttering the vowel /u/ or /i/, Ramachandran & Hubbard, 2001). Instead, it would seem more likely that the Bouba/Kiki effect constitutes a type of crossmodal correspondences in human perception (e.g., Marks, 1996; Walker, 1987), involving the matching of more rounded shapes with lower-pitched sounds whilst more angular shapes with higher-pitched sounds that has been observed at four months of age (Marks, 1987; Walker et al., 2010). Recent studies have successfully identified a

number of the critical visual features embedded in the patterns and acoustic cues in speech that drive the sound-shape correspondences (e.g., Chen, Huang, Woods, & Spence, 2016; Knoeferle, Li, Maggioni, & Spence, 2017; Westbury, Hollis, Sidhu, & Pexman, 2018). The Bouba/Kiki effect would therefore appear to be explainable on the basis of various crossmodal correspondences between specific visual and auditory features.

Chen et al. (2016) used radial frequency (RF) patterns (see Figure 1A for examples) to examine how people extract visual features and reach a consensual judgment in the Bouba/Kiki matching tasks. RF patterns are closed-contours with sinusoidal modulations along the circumference of a circle (Wilkinson, Wilson, & Habak, 1998). By manipulating the number and magnitude of sinusoidal modulations per circle, as well as the spikiness in terms of adding harmonic triangular wave forms on top of each sinusoidal modulation, these three RF-pattern features (i.e., Frequency, Amplitude, and Spikiness) can be changed in a stepwise manner. The result demonstrates that people are more likely to match an RF pattern to “Kiki” than to “Bouba” when the value of each factor – Frequency, Amplitude, and Spikiness – was increased. Hence, Chen et al. demonstrate that these three features in the RF patterns were valid predictors of people’s matching judgments, and the high consensus was driven by the cross-participant consistency in weighting each of the features. In contrast, the occurrence of minority responses can be attributed to people’s variability when sampling and evaluating features in the RF patterns.

Here we examined whether people can properly evaluate their confidence in sound-shape matchings in the Bouba/Kiki effect. We manipulated the critical features of RF patterns, giving rise to a continuous change of the consensual sound-shape matchings (Chen et al., 2016). Accordingly, we tested the correlations between individual’s subjective confidence and interpersonal consensus for a given pattern (e.g., Koriat, 2011). In addition, we compared the perceptual information on which the matching responses and confidence ratings were based.

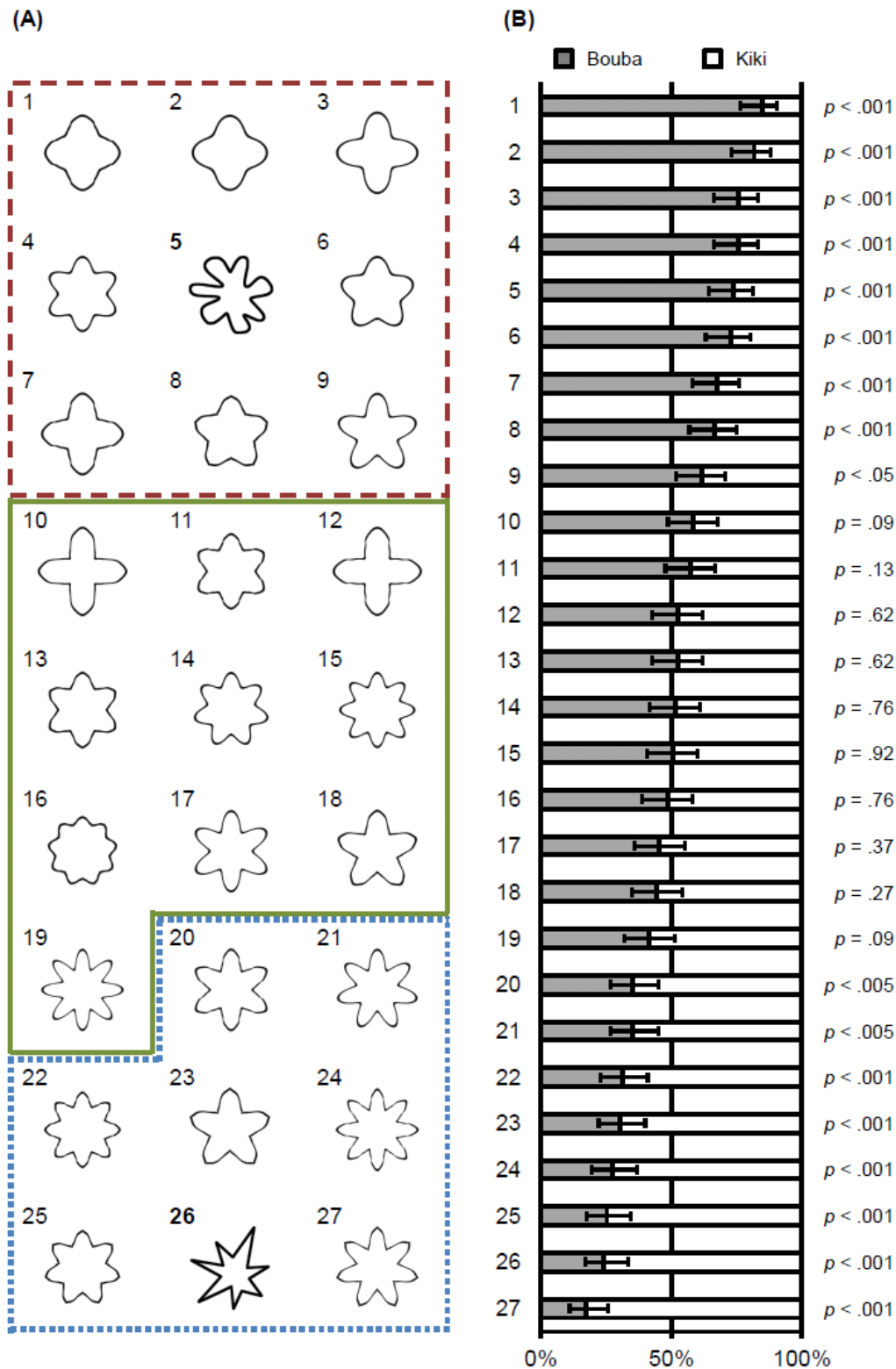


Figure 1. (A) The visual patterns used in the current study. These are RF patterns except number 5 and 26 (the two bold ones) which have been commonly used in previous studies testing the Bouba/Kiki effect. (B) The results (and the 95% confidence intervals) of the matching between each pattern to the nonsense speech sound “Bouba” or “Kiki”. The proportion of the “Bouba” and “Kiki” responses of each pattern were tested using the χ^2 test, and the p value is reported. According to the results, 9 patterns were predominantly matched to “Bouba” (in the red dashed column), 10 patterns were undetermined (in the green solid column), and the remaining 8 patterns were predominantly matched to “Kiki” (in the blue dotted column).

Methods

Participants

Ninety-nine participants (44 males, mean age: 23.9 years, range: 18-30 years) were recruited from Prolific Academic (<https://prolific.ac>) and took part in the study in exchange for £1 (UK Sterling). Prolific Academic's Country of Origin filter was used so as to only recruit those participants who were born and currently lived in the UK. Five additional participants were tested but excluded: three of whom failed to complete the experiment and the other two never reported a confidence in excess of 10%. All of the participants were naïve as to the purpose of the study. All of the participants gave their informed consent prior to the start of the experiment. The entire procedure was carried out in accordance with the Declaration of Helsinki and was approved by the ethical committee in Medical Sciences Inter Divisional Research Ethics Committee, University of Oxford (MSD-IDREC-C1-2014-141).

A hundred participants were planned to test before conducting the study in order to achieve adequate statistical power. An analysis using G*Power for chi-square test (version 3.1.9.2; Faul, Erdfelder, Lang, & Buchner, 2007) suggests that the power reaches 0.85 when testing 99 participants with type-I error (α) setting at the 0.05 level and a medium level of effect size (Cohen's $\omega = 0.3$).

Stimuli and Design

According to Chen et al.'s (2016) results, 25 RF patterns were selected spanning a spectrum from those that are consensually matched to “Bouba”, those that are undetermined, and those that are consensually matched to “Kiki” (see Supplementary Material for the functions to plot RF patterns). In addition, two typical patterns used in previous studies of the Bouba/Kiki effect were included (see 5 and 26 bolded in Figure 1A). Each RF pattern consisted of a black outline presented against a white background.

The auditory stimuli consisted of the spoken nonsense words “Bouba” and “Kiki” produced by a female native English speaker three times for each word with slightly different prosody (32 bit

mono; 44,100 Hz digitization). All six sound files were edited to the same length (400 ms) and their sound pressure level (in terms of the value of root mean square) were equalized. The experiment was conducted on the internet via the Adobe Flash based Xperiment software 2.0 (<https://www.xpt.mobi/>).

Procedure

Before starting the main experiment, the participants had to confirm that they could hear the sounds clearly (by typing in the three digits that were played out loud to them). After provided with a brief instruction about the task, the participants clicked a button in order to start the study when they were ready.

In each trial, a RF pattern was presented in the centre of the monitor, and the participants had to judge whether the words “Bouba” or “Kiki” (which they heard) provided a better match for the seen pattern – they had to choose one or the other in order to proceed. The order of presentation of “Bouba” and “Kiki” was randomized on a trial-by-trial basis. Next, the question “How confident are you regarding the match between the figure and the sound you chose?” was presented on the monitor, and the participants rated their confidence using a slider on the scale anchored at either end with the terms “not at all” on the left and “very much so” on the right. The pointer on the slider was initially hidden and only appeared when the participant moved the cursor over the line drawn between the two anchors. The pointer could be placed anywhere between the two ends of the scale, and the location was then converted into percentage values proportional to the two ends of the whole scale (“not at all” = 0%, and “very much so” = 100%).

The 25 RF patterns were presented in a completely randomized order. Subsequently, the two prototypical rounded and angular patterns (patterns 5 and 26) were tested in a random order as well. The study took around 5 min to complete.

Results

The agreement of participants' matching judgments for each pattern was assessed separately. That is, a chi-square test was used to determine whether the participants consistently judged a given pattern as better matching to "Bouba" or "Kiki", or else not different from chance level (50%; see Figure 1B). The 25 RF patterns can therefore be separated into three groups: Eight of them were significantly matched to "Bouba" ($X^2(1) \geq 5.34$, $ps < .05$, Cohen's $\omega \geq 0.23$), 10 of them were undetermined ($X^2(1) \leq 2.92$, $ps \geq .09$, Cohen's $\omega \leq 0.17$), and the remaining seven were significantly matched to "Kiki" ($X^2(1) \geq 8.50$, $ps < .005$, Cohen's $\omega \geq 0.29$). As expected, typical correspondences were observed between "Bouba" and the rounded pattern ($X^2(1) = 22.31$, $p < .001$, Cohen's $\omega = 0.48$), and between "Kiki" and the angular pattern ($X^2(1) = 26.27$, $p < .001$, Cohen's $\omega = 0.52$).

According to the above result, the participants' confidence ratings were analyzed in terms of two independent variables (see Figure 2): The *Category* factor was based on the results of the participants' matching judgments at the group level, so the 27 patterns were grouped as Bouba (9 patterns), undetermined (10 patterns), or Kiki (8 patterns). The *Judgment* factor was based on the matching judgments obtained from each participant, so the confidence ratings following "Bouba" or "Kiki" matching response were separated. In order to include both participants and patterns as random variables in the analysis, the linear mixed-effect model in the lme4 package (version 1.1-8, Bates et al., 2015) in R (version 3.2.5) was used with the model $\text{lmer}(\text{Rating} \sim 1 + \text{Category} * \text{Matching} + (1|\text{ID}) + (1|\text{Pattern}))$. Both *Category* ($t = -2.41$, $p < .05$) and *Judgment* ($t = -2.98$, $p < .005$) were significant, as was their interaction ($t = 5.87$, $p < .001$). The interaction was mainly driven by the higher confidence ratings when the "Kiki" pattern was matched to "Kiki" rather than to "Bouba" ($t = 4.25$, $p < .001$). However, the confidence ratings were not significantly different between the two judgments for the "Bouba" patterns ($t = -1.09$, $p = .28$), nor for the undetermined patterns ($t = 0.76$, $p = .45$).

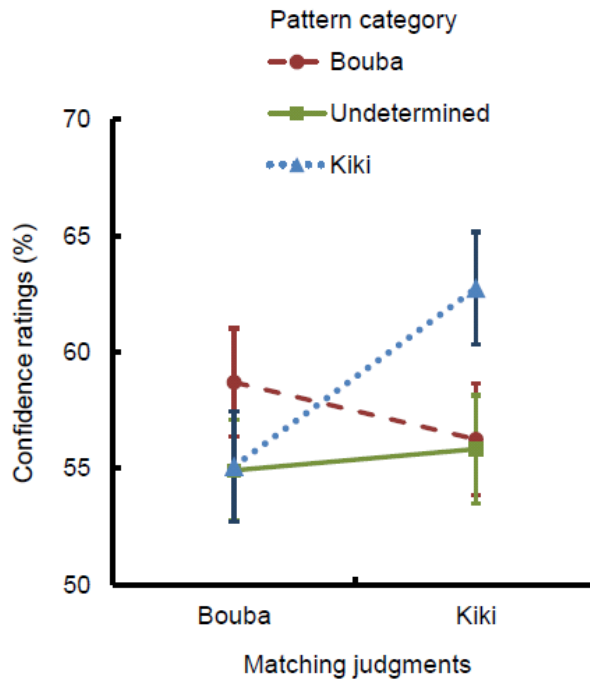


Figure 2. Mean confidence ratings (and the ± 1 standard error of the means) as a function of the three categories of patterns and the participant's own matching judgments for each pattern as either "Bouba" or "Kiki".

Figure 3A demonstrates the mean confidence ratings for each pattern as a function of the percentage of major responses (i.e., the response that more than 50% participants chose), irrespective of whether the major responses were "Bouba" or "Kiki". The significant correlation (Pearson $r(27) = 0.596$, $p < .001$) suggests that the participants were more confident of their answer for a given pattern with increasing matching consensus at the group level. Next, the average confidence rating for each pattern was taken from those "Bouba" and "Kiki" responses, separately. The correlation between confidence ratings and the percentage of "Bouba" responses was significant ($r(27) = 0.404$, $p < .05$; see Figure 3B). Nevertheless, when dividing the figures in terms of 50% "Bouba" responses, a significant correlation was observed only for those patterns with a consensus higher than 50% (i.e., the patterns with major responses to "Bouba", $r(15) = 0.546$, $p < .05$), rather than those below 50% (i.e., the patterns with minor responses to "Bouba", $r(12) = .023$, $p = .94$). Similarly, a significant correlation was observed between the confidence ratings and the percentage of "Kiki" responses ($r(27) = 0.580$, $p < .005$; see Figure 3C). Once again, we

observed that the significant correlation was mainly attributable to those patterns which were majorly matched to “Kiki” with consensus higher than 50% ($r(12) = 0.702, p < .05$), rather than those matched to “Kiki” with consensus below 50% ($r(15) = .009, p = .97$).

In order to further verify whether the confidence ratings and the matching consensus was positively correlated generally, or only in the major rather than in the minor responses, two types of linear-regression models were compared. Figure 4 demonstrates the re-plots of Figures 3(B) and (C) together with the trend lines based on each linear-regression model. Model 1 assumes that the confidence ratings (CR) increase with the percentage of “Bouba” or “KiKi” responses (PR). The equation can be expressed as $CR = a + b*PR$, where a is the intercept and b is the slope (Figure 4A and 4C). Model 2 assumes that the confidence ratings increase with the percentage of “Bouba” or “KiKi” responses only when it was higher than 50% (i.e., major responses), whereas the confidence ratings were a constant when the percentage of “Bouba” or “KiKi” responses was lower than 50% (i.e., minor responses; Figures 4B and 4D). Thus the equation can be expressed

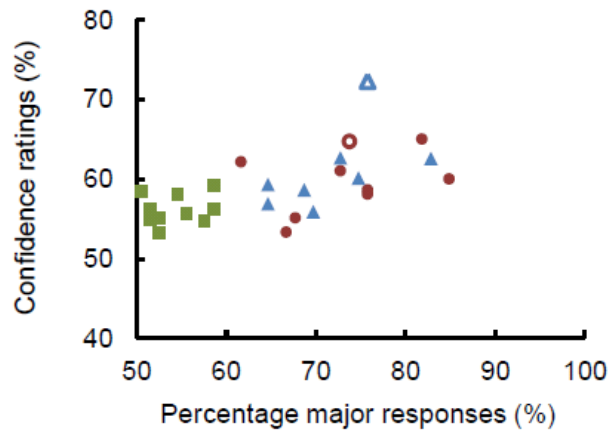
$$CR = a + b*PR, \text{ when } PR \geq 50\%,$$

$$CR = c, \text{ when } PR < 50\%.$$

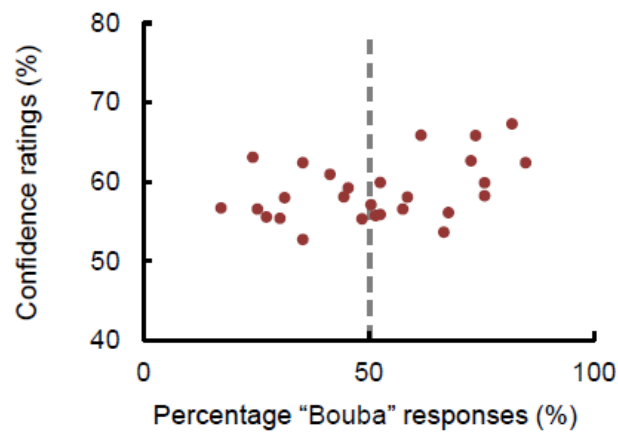
The fitted results and parameters were shown in Table 1. The goodness of fit of Models 1 and 2 were then compared using a nested F -test (Lu & Doshier, 2013). The results were significant for the “Kiki” responses, whilst not for the “Bouba” responses. The results suggest that, for the “Kiki” response, Model 2 did a better job than Model 1 at explaining the relationship between the confidence ratings and the matching consensus; that is, their positive correlation was only significant in the major rather than in the minor responses.¹ Nevertheless, for the “Bouba” response, both models explained the relationship similarly well, so the parsimonious suggestion is a general positive correlation between the confidence ratings and the matching consensus.

¹ When the outlier (percentage “Kiki” response = 75.8%, confidence rating = 75.2%) was excluded, both the results of correlation and model comparison for the “Kiki” response remained the same.

(A) Overall confidence ratings



(B) Confidence ratings of “Bouba” judgments



(C) Confidence ratings of “Kiki” judgments

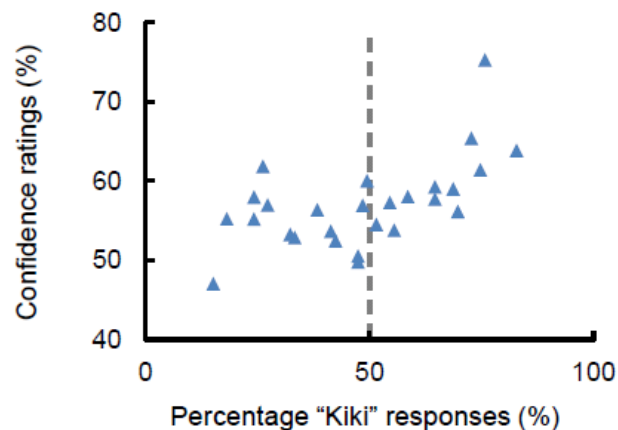


Figure 3. (A) Mean confidence ratings of each pattern as a function of the percentage of major responses (the response that more than 50% participants chose). The red circles represent the 9 “Bouba” patterns, the green squares represent the 10 undetermined patterns, and the blue triangles, the 8 “Kiki” patterns. The two open markers represent the rounded and angular patterns commonly used in previous studies of the Bouba/Kiki effect. (B) Mean confidence rating when the pattern was matched to “Bouba” as a function of the percentage of “Bouba” responses. (C) Mean confidence rating when the pattern was matched to “Kiki” as a function of the percentage of “Kiki” responses. The grey dashed line represents 50% of “Bouba” (or “Kiki”) responses.

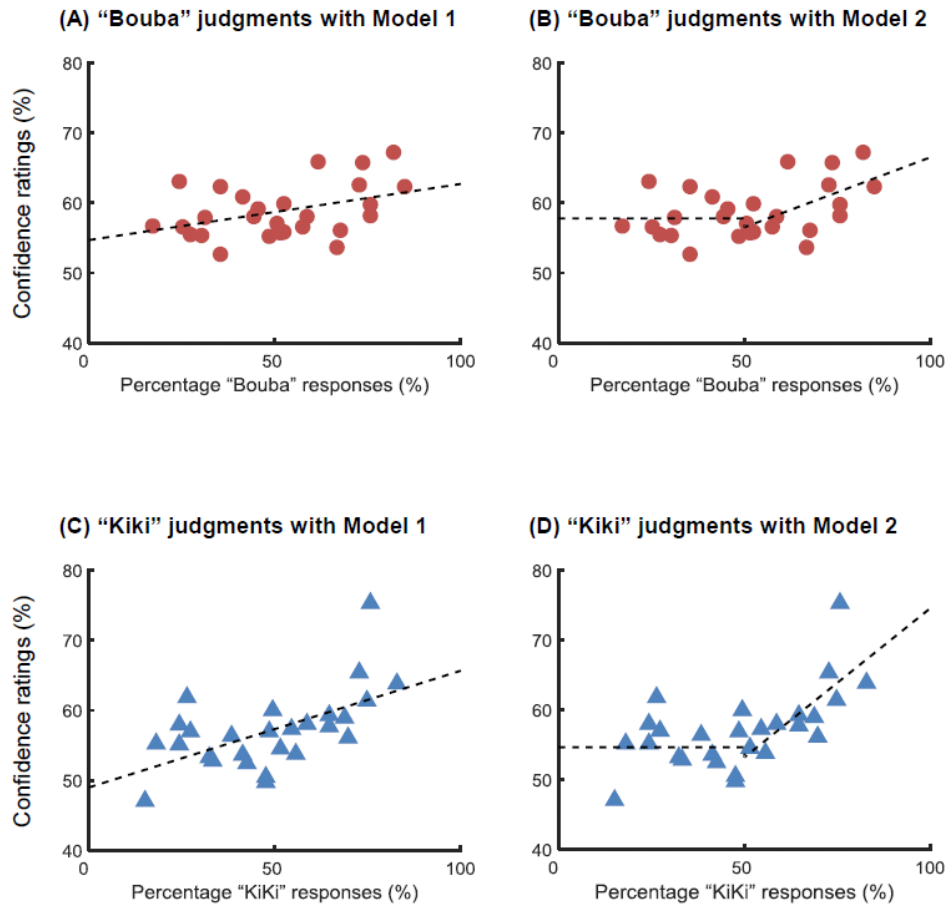


Figure 4. (A) The trend line of confidence rating under “Bouba” judgments using Model 1 which assumes that the confidence ratings increase with the percentage of “Bouba” responses. (B) The trend lines of confidence ratings under “Bouba” judgments using Model 2 which assumes that the confidence ratings increase with the percentage of “Bouba” responses only when it was larger than 50%, whereas the confidence ratings were a constant when the percentage of “Bouba” responses was lower than 50%. (C) The trend line of confidence rating under “Kiki” judgments using Model 1. (D) The trend lines of confidence ratings under “Kiki” judgments using Model 2.

Table 1. The parameters for each model and the results of model comparison

Response	Model	Parameters			SS_{res}	R^2	Model comparison	
		b	a	c			$F(1,24)$	p
Bouba	1	8.013	54.690	--	317.520	0.163	3.00	.096
	2	19.970	46.539	57.788	282.243	0.256		
Kiki	1	16.669	48.984	--	526.962	0.336	7.06	.014
	2	42.704	31.858	54.644	407.198	0.487		

Note: The nested F -test was used to compare the goodness of fit of the fuller model (Model 2) and the reduced model (Model 1). The equation can be expressed as $F(df_1, df_2) = [(r^2_{full} - r^2_{reduced})/df_1] / [(1 - r^2_{full})/df_2]$, where $df_1 = k_{full} - k_{reduced}$, $df_2 = N - k_{full}$, k_{full} is the number of parameters of the full model, $k_{reduced}$ is the number of parameters of the reduced model, and N is the number of predicted data points (Lu & Doshier, 2013).

The final analyses were designed to compare the information extracted from the RF patterns on which the participants' matching responses and confidence ratings were based. Specifically, we tested whether the three manipulated parameters of the 25 RF patterns – Frequency, Amplitude, and Spikiness – significantly and similarly predict the participants' matching responses and confidence ratings. Note that the matching response is a binary variable – either 0 (Bouba) or 1 (Kiki). The confidence ratings were therefore converted into a continuous variable ranging from 0 to 1 as well. The confidence ratings (in percentage) following the “Bouba” response was transformed using the equation $(1 - \text{confidence rating})/2$, whereas those following the “Kiki” response was transformed using the equation $(1 + \text{confidence rating})/2$. The values of confidence ratings for the “Bouba” responses then ranged between 0 to 0.5, with a lower value indicating a higher confidence in judging the pattern matched to “Bouba”. In contrast, the values of confidence ratings for the “Kiki” responses were between 0.5 to 1, with a higher value indicating a higher confidence in judging the pattern matched to “Kiki”. The logistic regressions in the lme4 package in R were used to fit the matching responses and confidence ratings separately using the maximum-likelihood method to reach the optimal coefficient for each factor. The parametric bootstrapping method was applied 1,000,000 times in order to derive the standard error (SE) for each coefficient, and therefore the 95% confidence interval (CI) can be calculated (see Table 2). The results indicate that all three factors were significant predictors for both matching responses and confidence ratings. Most critically, the coefficient of each factor was comparable in terms of the 95% CIs for matching responses and confidence ratings, suggesting that both responses were based on the same information of the patterns.

Table 2. The coefficient, SE, 95% confidence interval (CI = coefficient \pm 1.96*SE), z value, and p value ($\mu = 0$) for each factor in the logistic regression analysis.

Response	Factor	Coefficient	SE	95% confidence interval		z -values	p
				Upper	Lower		
Matching	Frequency	0.484	0.064	0.610	0.357	7.503	< .001
	Amplitude	0.074	0.014	0.101	0.047	5.317	< .001
	Spikiness	0.017	0.006	0.029	0.006	2.911	< .005
	Constant	-5.096	0.640	-3.843	-6.350	-7.966	< .001
Confidence rating	Frequency	0.461	0.037	0.533	0.388	12.494	< .001
	Amplitude	0.069	0.008	0.084	0.054	8.799	< .001
	Spikiness	0.017	0.003	0.023	0.010	5.090	< .001
	Constant	-4.868	0.372	-4.139	-5.597	-13.091	< .001

Note: The model of the logistic regression was `glmer(Response ~ 1 + Frequency + Amplitude + Spikiness + (1|ID) + (1|Pattern))`

Discussion

We examined people's introspection of self-confidence regarding their own sound-shape matching judgments in the Bouba/Kiki effect, one of the classic examples of crossmodal correspondences. The participants were more confident when their judgment was consistent with the major rather than with minor responses in the group, especially for those patterns predominantly matched to "Kiki". This more pronounced difference in confidence between the major and minor responses for "Kiki" than "Bouba" patterns is likely due to the fact that the mapping between "Kiki" and angular patterns is more reliable than that between "Bouba" and rounded patterns in the literature (Chen et al., 2016; Nielsen & Rendall, 2011; Woods, Spence, Butcher, & Deroy, 2013). We then went on to demonstrate that the mean confidence ratings were positively correlated with the interpersonal consensus across patterns. More critically, a general positive correlation was observed for the "Bouba" response, while the positive correlation was only significant when it was higher than 50% (i.e., major responses), rather than when it was lower than 50% (i.e., minor responses) for the "Kiki" response (see also Koriat, 2011). This positive correlation suggests that the critical information determining the Bouba/Kiki effect was highly agreed between participants,

and the participants were able to properly monitor their own processing in order to reach their matching judgments. On the other hand, the null correlation in the minor “Kiki” responses suggests that these responses should be driven by certain random uncertainties and/or variabilities during information processing (Juslin & Olsson, 1997; Koriat, 2011). Finally, the logit-regression analyses demonstrate that the three manipulated features of RF patterns – Frequency, Amplitude, and Spikiness – were able to predict the participants’ matching judgments as well as their confidence ratings in a similar way, suggesting that the decision in both responses is likely based on the same information.

The results therefore demonstrate that the Bouba/Kiki effect follows the consensuality principle: the positive correlation between the cross-participant consensus and individual’s confidence (Koriat, 2008, 2011). This result suggests that both the matching judgments and confidence ratings in the Bouba/Kiki effect should be underpinned by certain processes and knowledge that happen to be shared by the participants (Koriat, 2012). We suggest the common processes to be likely that the participants extracted critical features of the RF patterns, and then the information was used to complete the sound-shape matchings following the common knowledge. The common knowledge here are those crossmodal correspondence principles which are either intrinsically equipped or acquired based on the statistical regularities in the environment (Deroy & Spence, 2013, 2016; Spector, & Maurer, 2008, 2011; Spence, 2011). Interestingly, the results of logit-regression analyses suggest that the participants’ matching judgments and confidence ratings was predictable by the critical features with similar coefficient combinations. The remaining question therefore is whether the participants’ confidence ratings were directly determined by these featural cues just like their matching judgments (Gigerenzer, Hoffrage, & Kleinbölting, 1991). Or, alternatively, the confidence ratings might be based on the online feedback in terms of the accessibility and consistency of the crossmodal matchings between these featural cues (Koriat, 2012), giving rise to different levels of perceptual fluency that modulates people’s decision (e.g., Reber & Schwarz, 1999).

In summary, we highlight the metacognitive component to the Bouba/Kiki effect: People can well monitor their sound-shape matching performance which is underpinned by crossmodal correspondences, the knowledge that is commonly shared by participants. The current study provides a new method to explore novel unimodal or crossmodal correspondences and quantify their association strength using both matching consensus and confidence.

Acknowledgments

YCC and CS were supported by the Arts and Humanities Research Council (AHRC), Rethinking the Senses grant (AH/L007053/1). YCC is supported by Ministry of Science and Technology in Taiwan (MOST 107-2410-H-715 -001 -MY2). PCH is supported by Ministry of Science and Technology in Taiwan (NSC 102-2420-H-006-010-MY2 and MOST 105-2420-H-006-001-MY2).

References

- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ...
Grothendieck, G. (2015). *Linear Mixed-Effects Models using 'Eigen' and S4*. Retrieved from <https://cran.r-project.org/web/packages/lme4/lme4.pdf> (on 12/10/2015).
- Bremner, A., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K., & Spence, C. (2013). Bouba and Kiki in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition*, 126, 165-172.
- Chen, Y.-C., Huang, P.-C., Woods, A., & Spence, C. (2016). When “Bouba” equals “Kiki”: Cultural commonalities and cultural differences in sound-shape correspondences. *Scientific Reports*, 6:26681.
- Chen, Y.-C., & Spence, C. (2017). Assessing the role of the ‘unity effect’ on multisensory integration. *Frontiers in Psychology*, 8:445.
- Chen, Y.-C., Woods, A., & Spence, C. (2018). Sensation transference from plateware to food: The sounds and tastes of plates. *International Journal of Food Design*, 3, 41-62.
- Davis, R. (1961). The fitness of names to drawings. A cross-cultural study in Tanganyika. *British Journal of Psychology*, 52, 259-268.
- Deroy, O., & Spence, C. (2013). Why we are not all synesthetes (not even weakly so). *Psychonomic Bulletin & Review*, 20, 643-664.
- Deroy, O., & Spence, C. (2016). Crossmodal correspondences: Four challenges. *Multisensory Research*, 29, 29-48.
- Deroy, O., Spence, C., & Noppeney, U. (2016). Metacognition in multisensory perception. *Trends in Cognitive Sciences*, 20, 736-747.

- Faivre, N., Filevich, E., Solovey, G., Kühn, S., & Blanke, O. (2018). Behavioural, modeling, and electrophysiological evidence for supramodality in human metacognition. *Journal of Neuroscience*, 38, 263-277.
- Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98, 506-528.
- Holland, M. K., & Wertheimer, M. (1964). Some physiognomic aspects of naming, or, maluma and takete revisited. *Perceptual and Motor Skills*, 19, 111-117.
- Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review*, 104, 344-366.
- Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports*, 7:5562. doi:10.1038/s41598-017-05965-y.
- Köhler, W. (1929). *Gestalt psychology*. New York, NY: Liveright.
- Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York, NY: Liveright.
- Koriat, A. (2008). Subjective confidence in one's answers: The consensuality principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 945-959.
- Koriat, A. (2011). Subjective confidence in perceptual judgments: A test of the self-consistency model. *Journal of Experimental Psychology: General*, 140, 117-139.
- Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review*, 119, 80-113.
- Lu, Z.-L., & Doshier, B. (2013). *Visual psychophysics: From laboratory to theory*. Cambridge, MA: MIT Press.
- Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 384-394.
- Marks, L. E. (1996). On perceptual metaphors. *Metaphor and Symbolic Activity*, 11, 39-66.
- Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, 9, 316-322.

- Nielsen, A., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology*, 65, 115-124.
- Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, 114, 173-186.
- Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia – a window into perception, thought and language. *Journal of Consciousness Studies*, 8, 3-34.
- Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, 8, 338-342.
- Spector, F., & Maurer, D. (2008). The colour of Os: Naturally biased associations between shape and colour. *Perception*, 37, 841-847.
- Spector, F., & Maurer, D. (2011). The colors of the alphabet: naturally-biased associations between shape and color. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 484-495.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73, 971-995.
- Walker, R. (1987). The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. *Perception & Psychophysics*, 42, 491-502.
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, 21, 21-25.
- Westbury, C., Hollis, G., Sidhu, D. M., & Pexman, P. M. (2018). Weighing up the evidence for sound symbolism: Distributional properties predict cue strength. *Journal of Memory and Language*, 99, 122-150.
- Wilkinson, F., Wilson, H. R., & Habak, C. (1998). Detection and recognition of radial frequency patterns. *Vision Research*, 38, 3555-3568.
- Woods, A. T., Spence, C., Butcher, N., & Deroy, O. (2013). Fast lemons and sour boulders: Testing crossmodal correspondences using an internet-based testing methodology. *i-Perception*, 4, 365-379.